

# Interactive Visual Exploration of 3D Mass Spectrometry Imaging Data Using Hierarchical Stochastic Neighbor Embedding Reveals Spatiomolecular Structures at Full Data Resolution

Walid M. Abdelmoula,<sup>†,‡,§</sup> Nicola Pezzotti,<sup>‡</sup> Thomas Hölt,<sup>‡,§</sup> Jouke Dijkstra,<sup>†</sup> Anna Vilanova,<sup>‡</sup> Liam A. McDonnell,<sup>||,‡,▽</sup> and Boudewijn P. F. Lelieveldt<sup>\*,†,‡,▽</sup>

<sup>†</sup>Division of Image Processing, Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands

<sup>‡</sup>Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States

<sup>§</sup>Computational Biology Center and <sup>||</sup>Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands

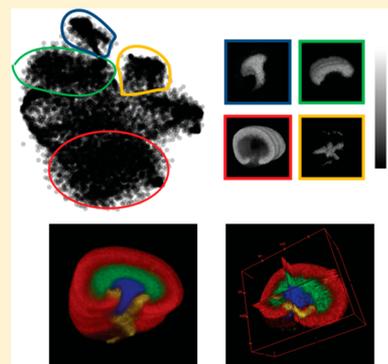
<sup>‡</sup>Computer Graphics and Visualization Group, Faculty of EEMCS, Delft University of Technology, 2628 CN Delft, The Netherlands

<sup>#</sup>Fondazione Pisana per la Scienza ONLUS, 56121 Pisa, Italy

## Supporting Information

**ABSTRACT:** Technological advances in mass spectrometry imaging (MSI) have contributed to growing interest in 3D MSI. However, the large size of 3D MSI data sets has made their efficient analysis and visualization and the identification of informative molecular patterns computationally challenging. Hierarchical stochastic neighbor embedding (HSNE), a nonlinear dimensionality reduction technique that aims at finding hierarchical and multiscale representations of large data sets, is a recent development that enables the analysis of millions of data points, with manageable time and memory complexities. We demonstrate that HSNE can be used to analyze large 3D MSI data sets at full mass spectral and spatial resolution. To benchmark the technique as well as demonstrate its broad applicability, we have analyzed a number of publicly available 3D MSI data sets, recorded from various biological systems and spanning different mass-spectrometry ionization techniques. We demonstrate that HSNE is able to rapidly identify regions of interest within these large high-dimensionality data sets as well as aid the identification of molecular ions that characterize these regions of interest; furthermore, through clearly separating measurement artifacts, the HSNE analysis exhibits a degree of robustness to measurement batch effects, spatially correlated noise, and mass spectral misalignment.

**KEYWORDS:** 3D MSI, data analysis, segmentation, proteomics, nonlinear dimensionality reduction, t-SNE, HSNE



## 1. INTRODUCTION

Mass spectrometry imaging (MSI) is a promising technology for many life science and biomedical applications.<sup>1–3</sup> MSI can provide the spatial distribution of hundreds of biomolecules directly from tissue. Typically, a thin tissue section is analyzed, pixel-by-pixel, in a predefined 2D raster. Matrix-assisted laser desorption ionization (MALDI),<sup>4,5</sup> secondary ion mass spectrometry (SIMS),<sup>6</sup> and desorption electrospray ionization (DESI)<sup>7</sup> are among the most common ionization methods. MALDI can be used to analyze a diverse range of molecular classes just by changing the tissue preparation method, DESI is able to provide molecular information about lipids without any tissue preparation, and SIMS provides very high spatial resolution capabilities also without any tissue preparation.

MSI may also be performed in three dimensions, most often by 2D MSI analysis of sequential tissue sections followed by their coregistration into a 3D volume.<sup>8–13</sup> It has been shown that 3D MSI data can be integrated with in vivo imaging

modalities such as magnetic resonance imaging (MRI),<sup>8,14,15</sup> fluorescence microscopy,<sup>16</sup>  $\mu$ -CT,<sup>17</sup> and positron emission tomography (PET).<sup>18</sup> This integration is not only useful from the biological viewpoint<sup>19,20</sup> but also important for the coregistration process that is required to construct the 3D MSI data sets.<sup>21</sup> In vivo imaging modalities preserve the geometrical entity of the tissue volume and thus provide a reference that may be used to construct and visualize the 3D molecular maps.

For MALDI- and DESI-based experiments, 3D MSI is essentially the merging of the 2D MSI data sets from a stack of serial tissue sections.<sup>11</sup> Recent technological advances allow 3D MSI to be acquired in a reasonable time frame.<sup>22,23</sup> Each voxel's mass spectrum is represented by three spatial coordinates ( $x, y, z$ ), and the 3D MSI data set can contain millions of voxels

**Received:** October 9, 2017

**Published:** February 12, 2018

Table 1. Summary of the 3D MSI Data Sets And Their Computational Processing Time Using HSNE

data set	preservation	mass range (kDa)	no. tissue sections; (tissue thickness $\mu\text{m}$ )	spatial resolution ( $\mu\text{m}$ )	data set size (no. voxels $\times$ no. $m/z$ features)	HSNE running time (min)
3D DESI-MSI colorectal carcinoma	fresh frozen	0.2–1.05	26; (10)	100	148 044 $\times$ 8073	$\sim$ 10
3D MALDI-MSI mouse kidney	PAXgene	2–20	73; (3.5)	50	1 362 830 $\times$ 7680	$\sim$ 43
3D MALDI-MSI mouse pancreas	PAXgene	1.6–15	29; (5)	60	497 255 $\times$ 13 312	$>$ 25
3D MALDI-MSI OSCC	fresh frozen	2–20	58; (10)	60	825 558 $\times$ 7680	$\sim$ 30
3D MALDI-MSI atherosclerotic plaques	fresh frozen	$<$ 1	5; (10)	100	10 185 $\times$ 20	$\sim$ 5

and mass spectra. This hyper-dimensional 3D MSI data provides rich molecular information on high chemical specificity across the entire tissue volume but poses computational challenges to efficiently analyze, visualize, and identify informative patterns.<sup>24</sup> Currently, there are needs and ongoing interests of developing computational methods to tackle these challenges.<sup>10,21,25,26</sup>

Dimensionality reduction is a well-established component for handling and analyzing high-dimensional data.<sup>27–29</sup> It seeks to represent the high-dimensional data in a lower dimensional space, to facilitate efficient visualization, classification, and clustering.<sup>27</sup> Common linear dimensionality reduction algorithms such as principal component analysis (PCA)<sup>30</sup> and non-negative matrix factorization (NNFM)<sup>31</sup> have been widely used for analyzing 2D MSI data sets<sup>32,33</sup> and have also been applied to 3D MSI data sets.<sup>16,34,35</sup> Nevertheless, their inherent linearity constraints mean that the analyses will be dominated by the major differences in the data sets, for example, between different cell types within the tissue volume.<sup>36</sup>

State-of-the-art nonlinear dimensionality reduction is a family of algorithms inherited from Stochastic Neighbor Embedding (SNE).<sup>37</sup> The hallmark of these algorithms is their ability to preserve local structures of high-dimensional data in a low map representation. t-Distributed stochastic neighbor embedding (t-SNE) enables the visualization of high-dimensional nonlinear data by alleviating the crowding problem of SNE and thus is able to visualize high-dimensional data in a single map representation.<sup>36,38–40</sup> Fonville et al.<sup>41</sup> and Abdelmoula et al.<sup>42</sup> have highlighted the superiority of t-SNE for analyzing 2D MSI data sets. Nevertheless, the quadratic computational complexity of t-SNE has limited its practical applicability to data sets of up to a few thousand data points.<sup>43</sup> The Barnes–Hut SNE (BH-SNE), an accelerated version of the t-SNE, has subsequently been shown to handle larger data sets of up to a few hundred thousand data points with a computational complexity of  $O(N \log N)$  where  $N$  is the number of data points.<sup>40,43–46</sup>

With current increases in data size, in which data sets can contain up to millions of data points, BH-SNE also becomes impractical.<sup>47,48</sup> On such large data scales BH-SNE becomes computationally intractable, and the interpretation of the final crowded embedding is nontrivial as it visualizes millions of data points in a single 2D or 3D scatter plot.<sup>48</sup> Recent progress has been made by Pezzotti and coworkers, in which the hierarchical stochastic neighbor embedding (HSNE) is used to create a hierarchical representation of the nonlinear data, allowing scalable exploration of the high-dimensional space in a low-dimensional space by constructing 2D embeddings that contain a few hundred data points.<sup>47</sup>

The HSNE algorithm aims at visualizing meaningful landmarks that represent sets of high-dimensional data points and has been shown to preserve rare but potentially disease-related clusters.<sup>48</sup> The HSNE technique is based on the concept “Overview-First, Details-on-Demand”.<sup>49</sup> This means that on the higher, coarser, hierarchical scale the resultant embedding shows dominant data structures (i.e., an overview). Then, a more detailed information can be visualized by computing a new embedding at the subsequent finer hierarchical scale using a selection of landmarks of dominant structures from the higher scale and so on. Eventually, this interactive hierarchical scheme helps the user to iteratively refine the visualized information and find informative structures on different scales, while keeping both memory and computational complexities manageable. This is because the landmarks used on a finer scale are a subset of the previous, recomputed, coarser scale. For more detailed information about the HSNE algorithm, we refer to Pezzotti et al.<sup>47</sup>

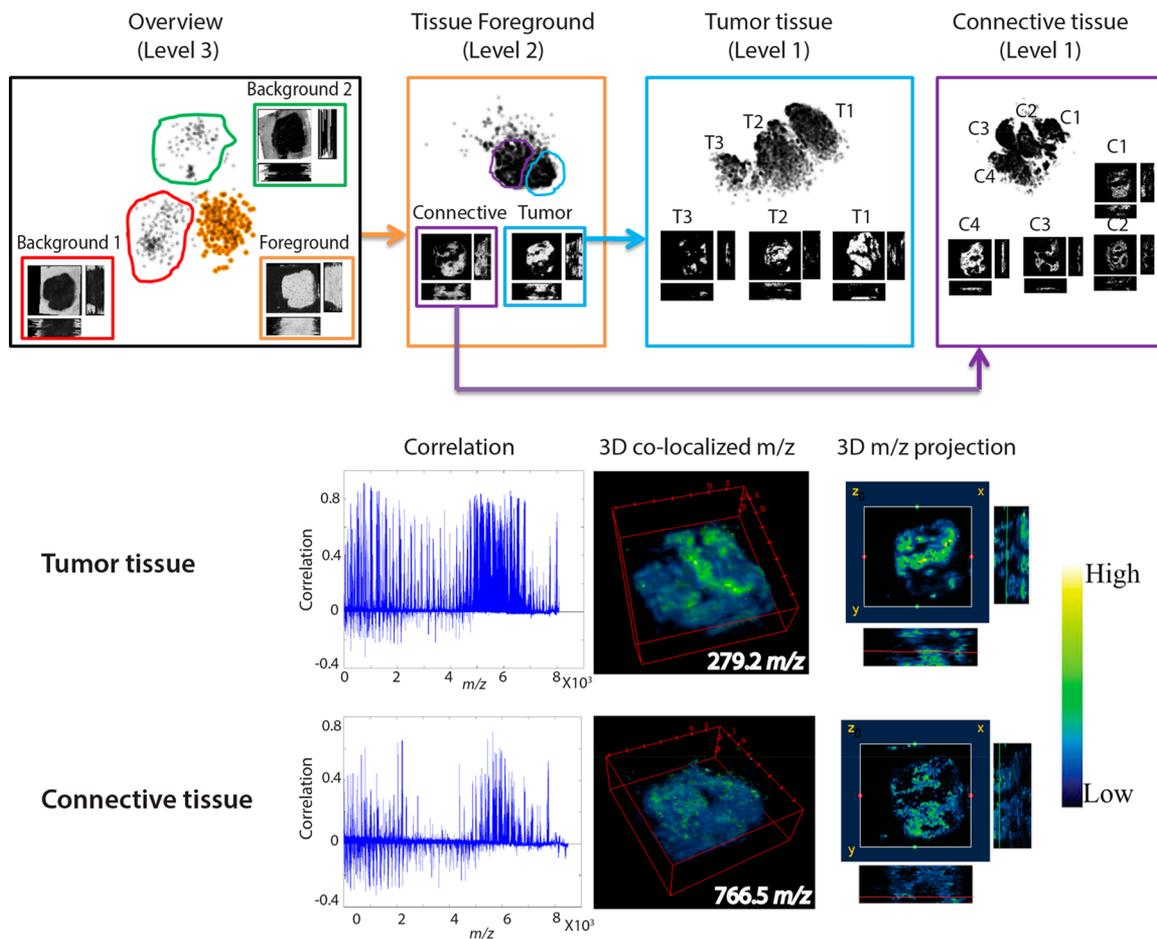
Recently, Oetjen et al. published a set of benchmark 3D MSI data sets, which were acquired using different ionization techniques and collected from different biological systems, namely, murine kidney, murine pancreas, human colorectal adenocarcinoma, and human oral squamous cell carcinoma.<sup>10</sup> This data is publicly available and can be downloaded from the *GigaScience GigaDB* repository.<sup>10</sup> Patterson and coworkers have also recently published 3D MSI data set of lipids in human carotid atherosclerotic plaque.<sup>50</sup> 3D MSI data sets may easily consist of millions of voxels, with thousands of spectral features per voxel. Until now, processing such data sets with t-SNE type approaches was computationally not feasible.

We investigate whether HSNE can be deployed to analyze complete 3D MSI data sets, at full resolution, to reveal tissue-specific spectral signatures at dense spectral and spatial resolution. To this end, we present a framework that consists of (a) dimensionality reduction and data visualization using HSNE, (b) a method to derive 3D maps from selected structures in the HSNE embeddings, and (c) a method to identify tissue-specific  $m/z$  features using the 3D spatial correlations between the 3D HSNE maps and the original 3D MSI data. We validate the proposed approach in a variety of previously published 3D MSI data sets from different biological systems.

## 2. MATERIALS AND METHODS

### 2.1. Experimental Data Sets

The 3D MSI data sets used in this study are from a previous study of Oetjen and publicly available for download from the *GigaScience GigaDB* repository.<sup>10</sup> These data sets were acquired by different MSI ionization methods and were collected from five different biological systems, namely, mouse kidney, mouse



**Figure 1.** Hierarchical analysis of 3D DESI-MSI of colorectal carcinoma data set using the HSNE reveals structural patterns at different hierarchical scales. The overview embedding represents the coarsest level in which generic dominant structures are revealed, namely: background and foreground tissue. Detailed embedding on the tissue foreground reveals two major structures that represent colorectal cancer and connective tissues. At the finest embedding level, more structures are uncovered within each of the colorectal cancer and muscle tissues. The Pearson correlation distribution between HSNE segmentation maps at Level 2 and all of the spectra is presented for cancer and muscle tissue, showing the most localized  $m/z$  feature in both tissue classes.

pancreas, human colorectal adenocarcinoma, cultured interacting microbial colonies, and human oral squamous cell carcinoma. A brief description of each data set is given in Table 1.

## 2.2. HSNE on 3D MSI Data

Each 3D MSI data set was organized in a matrix format  $M_{n \times f}$  in which  $n$  is the number of spectra (i.e., number of voxels) and  $f$  is the number of  $m/z$  features in each spectrum. The HSNE algorithm was applied to  $M_{n \times f}$  to find a hierarchical and multiscale representation,  $L$ . The term  $L^s$  refers to the set of low-dimensional landmarks that represent the data set on scale  $s$ . The first scale  $L^1$  represents the original data points of  $M$ , and landmarks of higher scales are subsets of previous scales ( $L^s \subset L^{s-1}$ ), in which the landmarks are automatically selected to represent a set of data points.

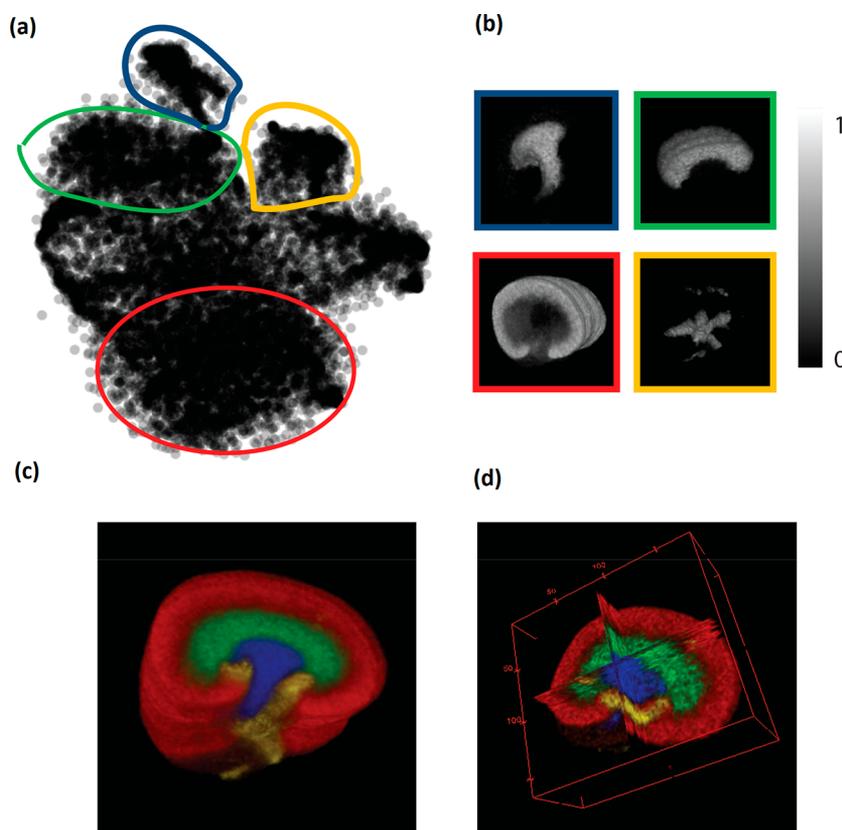
The HSNE algorithm starts at  $L^1$  by defining a Finite Markov Chain (FMC) that works as similarity matrix  $P_1$  for the data points with linear memory complexity and computational complexity  $O(n \log(n))$ . Landmarks are selected by computing the stationary distribution of the FMC and selecting the data points whose stationary value is higher than a given threshold, and this step has a computational complexity of  $O(|L^s|)$ . The “area of influence” of landmarks in  $L^2$  on landmarks in  $L^1$  is also

computed, which is a probability function that encodes the relatedness of the landmarks in  $L^2$  with the data in  $L^1$ . The calculation of the area of influence has a computational complexity of  $O(|L^{s-1}|)$  and a memory complexity that grows linearly with the size  $L^2$ . Finally, the similarity matrix  $P_2$  between the landmarks in  $L^2$  is computed as the pairwise overlap of the corresponding areas of influence.

To construct a lower dimensional representation (2D) of the landmarks in  $L^2$ , the t-SNE algorithm is applied using  $P_2$  as input instead of the Euclidean distances between the original data points in  $L^1$ . The power of the HSNE algorithm is to further iterate this process, in which the process above is repeated using  $P_2$  as FMC for landmarks in  $L^2$  and computing the next hierarchical scale  $L^3$  and so on.

The application of t-SNE to the landmarks at level  $L^s$ , using the similarity matrix  $P_s$  as input, reveals clusters of landmarks; the hierarchical nature of the landmarks mean that these clusters represent larger structures in the high-dimensional data (and in which the hierarchical level determines the scale of the data structures revealed by the t-SNE analysis).

The steps described above result in a hierarchical representation of the data, in which landmarks have been automatically detected as data points that are representative for a group of neighbors in the data space. The t-SNE maps of the



**Figure 2.** Analysis of 3D MALDI-MSI data of a mouse kidney using the HSNE: (a) HSNE scatter plot showing the spectral similarities as landmarks in a low-dimensional representation and (b) HSNE spatial structures based on the landmarks selection in panel a. The identified four anatomical structures with distinct spectral signatures were merged into a single 3D image (c,d) representing: renal cortex (red), renal medulla (green), renal pelvis (blue), and surrounding of renal pelvis (yellow). The multiorthoslice view in panel d allows in-depth visualization of the identified features.

landmarks at any level of the data hierarchy can be explored interactively by manually annotating a cluster in the t-SNE maps and drilling into the data underlying the landmarks.

Heterogeneity within the larger scale structures can be revealed by first selecting the data within the cluster (given by the area of influence of each landmark contained in the cluster) and creating embeddings at a lower hierarchical level. In this manner HSNE enables a hierarchical exploration of very high-dimensionality data. It should therefore be noted that during generation of the hierarchy, landmarks are selected automatically from the data; during the exploration, subsets of landmarks are selected in this case by manual drawing of clusters of landmarks and subsequently drilling into the data in the level below. For more details of HSNE and t-SNE, we refer the interested reader to the original papers.<sup>38,47</sup> In addition, the source code of the HSNE algorithm has recently been released and is publicly available.<sup>48</sup>

### 2.3. HSNE Spatial Segmentation Maps

Every landmark in the HSNE embedding holds probability values representing the likelihood, for each of the original high-dimensional data points, of belonging to that landmark. The landmarks are located in the HSNE embedding based on their mass spectral similarities. This means that mass spectrometrically similar landmarks cluster together, whereas dissimilar landmarks are located further apart, frequently with clear boundaries between clusters. Here we manually selected clusters that could also be automated using a density-based partitioning.<sup>36</sup>

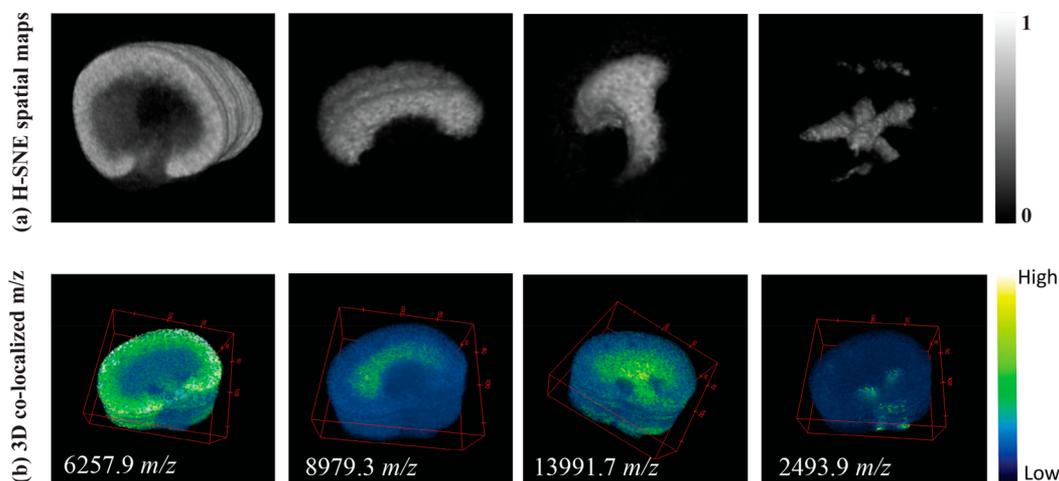
Once a cluster of landmarks has been selected, a spatially resolved HSNE segmentation map can be constructed. The HSNE segmentation map is a 3D gray-scale image with intensity values ranging between  $[0,1]$ ; these reflect the probability of the voxel belonging to the selected landmarks. Voxels of high probability values have a similar mass spectrum to one of the selected landmarks, whereas voxels of low probability values are not represented by that particular selection of landmarks; their similarities are encoded by other landmarks in the HSNE scatter space.

The HSNE spatial segmentation maps reveal multiscale spatial structures, and the spatial scale depends on the hierarchical level of the HSNE embedding from which the spatial structures were originally reconstructed. Therefore, finer HSNE spatial structures are typically constructed from landmarks in the HSNE embedding on a finer hierarchical scale and so on.

Eventually, an HSNE spatial segmentation map depicts a region of interest that shares similar mass spectral characteristics. Unlike hard clustering techniques such as  $k$ -means,<sup>51</sup> the HSNE spatial segmentation map can be considered as a fuzzy-like cluster<sup>52</sup> in which each data point in the entire data set holds a probability of belonging to the cluster.

### 2.4. Spatial Correlations and Corresponding $m/z$ Colocalization

The HSNE segmentation map reflects a specific structure in the 3D MSI data, which can be used to identify the molecular ions that exhibit similar spatial distributions. A colocalized  $m/z$  feature is highly expressed in the structure highlighted by the



**Figure 3.** Visualization of the most colocalized 3D  $m/z$  features with respect to the associated HSNE spatial segmentation maps of the 3D MALDI-MSI mouse kidney data set.

HSNE segmentation map and lowly expressed elsewhere. Colocalized  $m/z$  features can be identified by first calculating the Pearson correlation between  $m/z$  images and the HSNE segmentation map and then determining those that achieve significant correlation score ( $p$  value  $<0.05$ ). It is possible to identify more than one colocalized  $m/z$  feature; however, in this manuscript and for presentation simplicity we opted to visualize only the highest colocalized features.

### 3. RESULTS

#### 3.1. 3D DESI-MSI of Colorectal Carcinoma

The low-dimensional representation generated by HSNE of the 3D DESI-MSI data set of colorectal carcinoma is shown in Figure 1. The HSNE scatter plots show patterns of landmarks that were projected, at different hierarchical levels, based on their similarities in the high-dimensional space. Figure 1 visualizes the hierarchical representation at three embedding levels, ranging from overview to detailed visualization. Level 3 represents the overview embedding, which visualizes the more global patterns in the data set and separates the tissue foreground from the background. Two clusters representing the background were detected, which is presumed to reflect the heterogeneous nature of the background noise in the original high-dimensionality data. To drill-in to more detailed structures the tissue foreground cluster was selected and a new embedding was constructed at the next level. The level 2 embedding of the tissue foreground revealed two new structures, representing colorectal cancer and connective tissue. This is in agreement with Oetjen et al., who reported two main tissue types (tumor and connective tissue) based on histopathological examination of the tissues.<sup>10</sup> Supplementary Figure S1 demonstrates the close similarity of demarcating tumor from connective tissue in the histological images and the HSNE segmentation maps of level 2. When the cancer and connective tissues were separately subjected to HSNE at the finest hierarchical level, level 1, new structural features were revealed in the HSNE space and associated 3D data volume (Figure 1). Figure 1 also shows the Pearson correlation distributions between the HSNE segmentation maps at embedding Level 2 and all of the voxel associated mass spectra as well as the distributions of the ions with highest correlation for cancer and connective tissues, respectively.

The HSNE algorithm automatically constructed the three hierarchical levels in 10 min on a PC workstation with a 3.5 GHz Intel Xeon processor and 128 GB memory, resulting in the overview embedding. The subsequent, more detailed embeddings required 2 min or less to be visualized based on landmark selection at the previous embedding level.

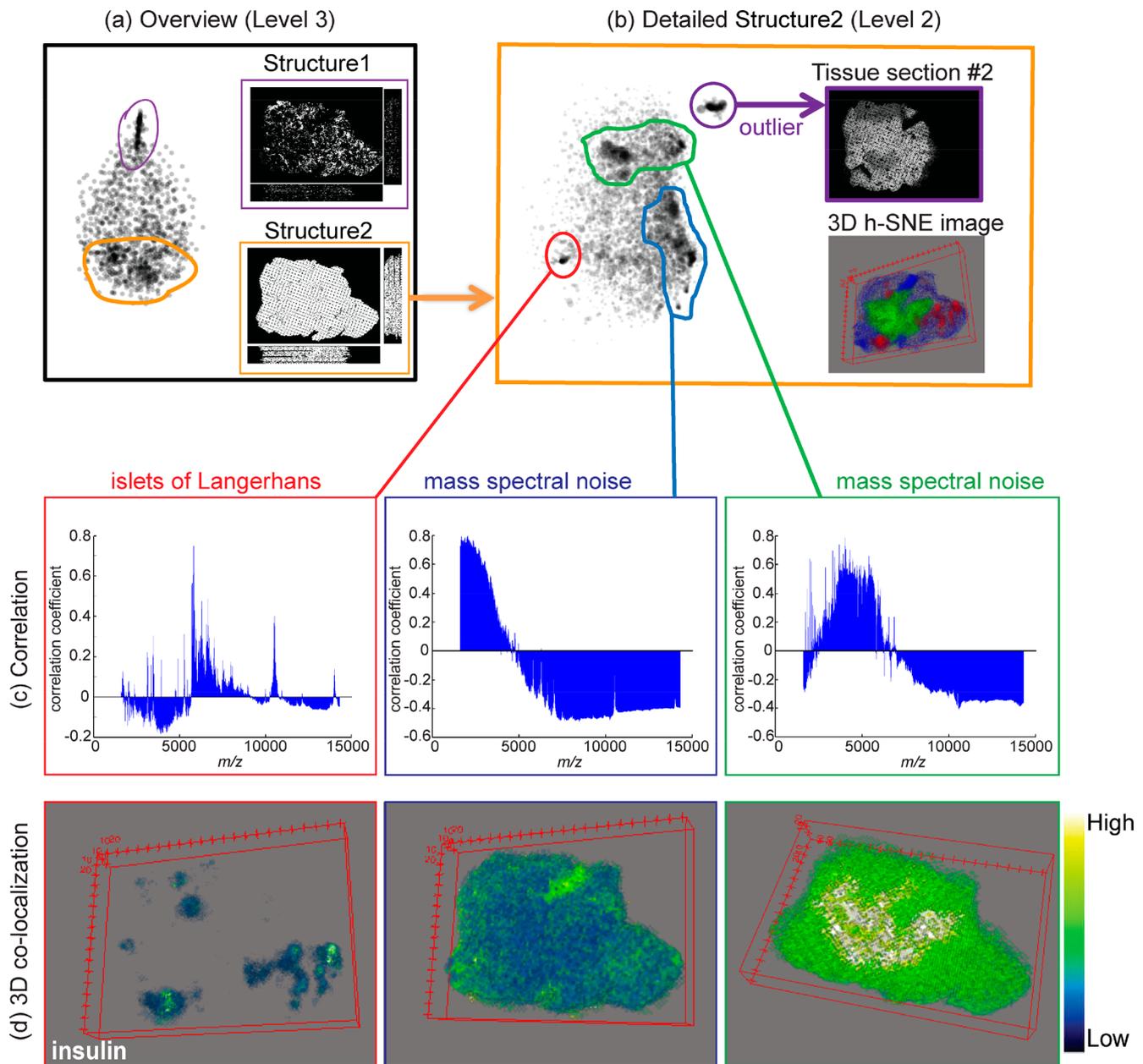
#### 3.2. 3D MALDI-MSI of Mouse Kidney

The 40 GB 3D MALDI-MSI data set of the mouse kidney was analyzed using the HSNE pipeline, and the resulting structural patterns are shown in Figure 2. The HSNE algorithm automatically constructed four hierarchical levels from this large data set, which were computed in  $\sim 43$  min on the same PC referred to above. For ease of visualization the structures at hierarchical embedding level 2 were selected and are presented in Figure 2a in the HSNE space; Figure 2b shows the associated 3D HSNE segmentation images (which displays each voxel's probability of belonging to the selected cluster of landmarks). In agreement with Trede et al.<sup>9</sup> who previously processed this mouse kidney data set at reduced size, four main anatomical structures in the mouse kidney were identified, but in this instance the calculation was performed on the full data set and revealed finer spatial detail. Figure 2c,d shows the four regions as false-color 3D volumes, specifically the renal cortex (red), renal medulla (green), renal pelvis (blue), and the surrounding of the renal pelvis (yellow). Of note, the landmarks not selected in the level 2 embedding represent noise-related structures; see Supplementary Figure S2.

The 3D structures corresponding to the tissue clusters identified by the HSNE were then used to identify which molecular ions exhibited highly correlated colocalization. The Pearson correlation between the 3D HSNE spatial clusters and the spectral images were calculated (see Supplementary Figure S3), and the colocalized  $m/z$  features with the highest correlations were identified and are shown in Figure 3. Supplementary Figure S4 shows the 3D projections of these colocalized ion features.

#### 3.3. 3D MALDI-MSI of Mouse Pancreas

The 3D MALDI-MSI data set of the mouse pancreas was analyzed using the HSNE pipeline, and the resulting structural patterns are shown in Figure 4. Three hierarchical embedding levels were automatically constructed, and the HSNE running time is reported in Table 1. The coarser embedding at level 3

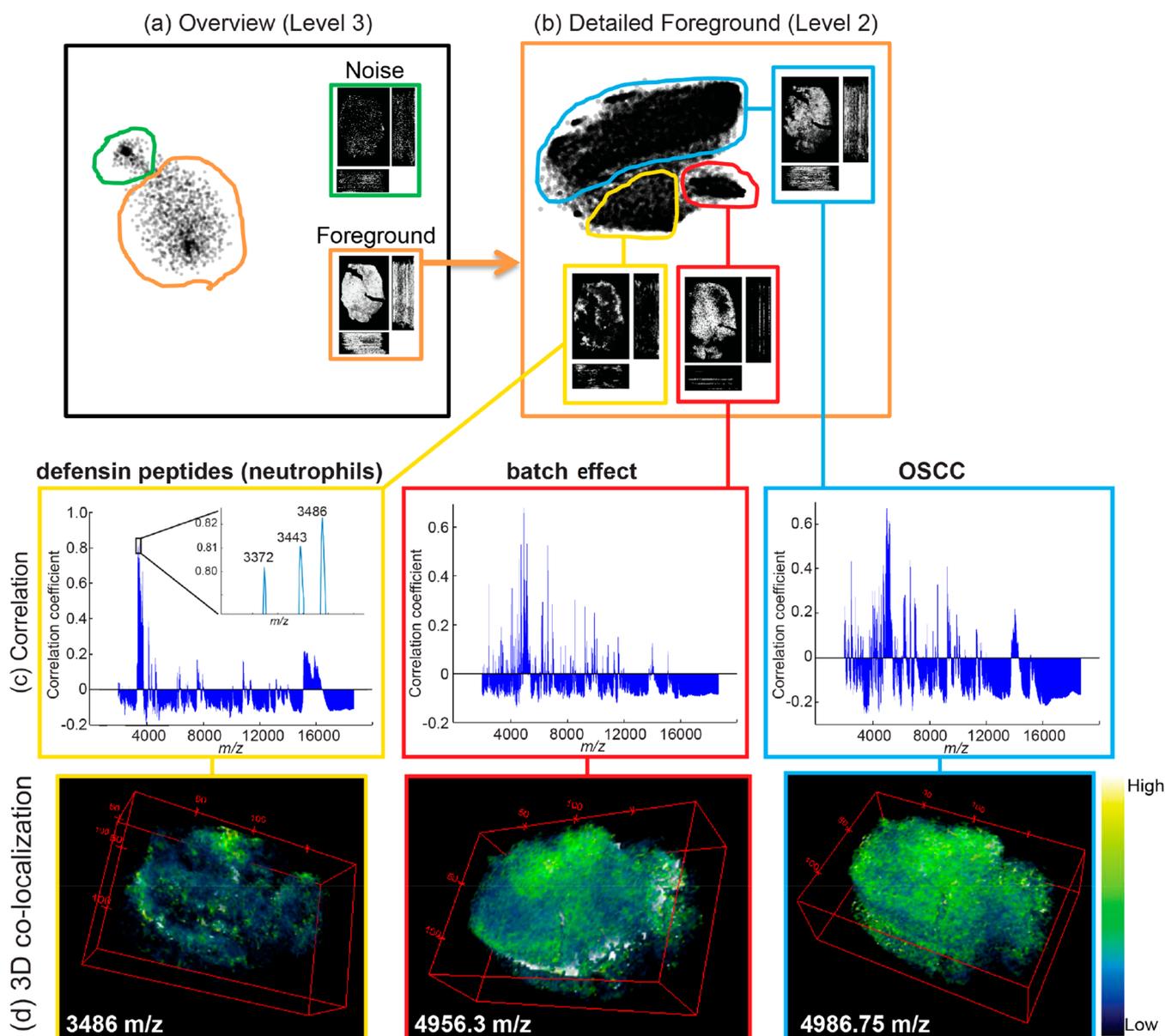


**Figure 4.** Analysis of 3D MALDI-MSI of mouse pancreas data set using the HSNE reveals structural patterns at different hierarchical scales. The detailed embedding at level 2 reveals three spectrally distinct clusters given in panel b and colored red, green, and blue. The spatial correlation between each of the clusters identified in panel b and the spectral information was computed (c), and the highest localized  $m/z$  features were identified (d). The  $m/z$  value of 5805.54, which is colocalized with the red cluster given in panel b, was previously identified as insulin.

differentiated between noise and two tissue structures, termed structure 1 and structure 2 (Figure 4a). No additional structural information was revealed within structure 1 at subsequent embedding levels, and so correlation analysis was computed at this level and revealed spatially correlated mass spectral noise; see Supplementary Figure S5. For tissue structure 2 (h-SNE level 3) a more detailed embedding at the next level revealed a highly structured data space (Figure 4b). Close examination of the HSNE map revealed the data structures distinguished highly localized regions characterized by distinct molecular profiles (red cluster), outlier tissue sections (purple cluster), and spatially correlated mass spectral noise (blue and green clusters). Each of these structures is defined by distinct mass spectral profiles and 3D spatial distributions (Figure 4c,d, respectively). The protein ion that displayed the greatest

colocalization with the red cluster,  $m/z$  5805.54, was reported by Oetjen et al.<sup>10</sup> in the original benchmark 3D MSI data sets paper as insulin. Insulin is produced by the beta cells in islets of Langerhans, highly localized endocrine tissue in the pancreas, which are known to exhibit very distinct spatial and molecular profiles. HSNE enabled highly localized features to be rapidly identified in a large 3D MSI data set, even when that data set contained outlier tissue sections and significant spatially correlated noise.

Drilling-in to the subsequent finer hierarchical level (level 1), no new structures were identified and therefore we based our results on the two embedding levels presented in Figure 4a,b.



**Figure 5.** Analysis of 3D MALDI-MSI of human oral squamous cell carcinoma data set using the HSNE reveals structural patterns on different hierarchical scales. The correlation analysis (c) allows us to identify the most colocalized  $m/z$  features (d) with the HSNE spatial structures (b).

### 3.4. 3D MALDI-MSI of Human Oral Squamous Cell Carcinoma (OSCC)

The 3D MALDI-MSI data set of OSCC was analyzed using the HSNE pipeline, and the resulting structural patterns are shown in Figure 5. Three hierarchical embedding levels were automatically constructed, which were computed in less than half an hour on the same PC referred to above (Table 1). The coarser embedding at level 3 distinguished two dominant patterns, namely, noise and tissue structure (Figure 5a). A more detailed embedding of the tissue foreground was constructed at hierarchical level 2 (Figure 5b) and revealed three structures. The correlation distribution between the 3D HSNE cluster maps and the 3D MSI data is shown in Figure 5c. The molecular ions with the highest colocalization metrics were identified, and their 3D distributions are shown in Figure 5d. The peptide ions at  $m/z$  3486, 3443, and 3372 were strongly colocalized with the yellow HSNE cluster and were previously reported by Oetjen et al.<sup>10</sup> as defensins HNP1–3, peptides

produced by neutrophils (HNP refers to Human Neutrophil Peptide). The mass spectra associated with the red and blue clusters were similar, consisting of the same peptide and protein ions but with different relative intensities. Close examination of the 3D distributions revealed that the red cluster was characterized by a batch effect, in which a number of tissue sections (tissue section numbers 31, 32, and 33) were characterized by very intense thymosin  $\beta_4$  signals, which can be observed as white banding in Figure 4d. Supplementary Figure S6a shows a comparison of the average mass spectra from tissue section number 1 and tissue section number 31, one of those exhibiting a strong batch effect, for the thymosin  $\beta_4$  signals. Close examination of the spectra also indicated small mass shifts between the spectra; the HSNE algorithm does not include a mass spectral alignment step, and so such misalignment of spectra would be interpreted as different molecular signatures and their separation into separate clusters. Supplementary Figure S6b shows the batch-affected tissue

sections are localized to specific regions of the 3D MSI data set. Nevertheless, as with the 3D MSI data set of pancreas, HSNE enabled meaningful conclusions to be rapidly extracted from a large 3D MSI data set, even if it contained batch effects (intense mass spectral peaks and mass spectral misalignment).

### 3.5. 3D MALDI-MSI of Human Atherosclerotic Plaques

The 3D MALDI-MSI data set of human atherosclerotic plaques was analyzed by the HSNE pipeline, and the resulting structural patterns are shown in [Supplementary Figure S7](#). On the basis of the data distribution, two hierarchical embedding levels were automatically constructed. The coarser embedding at level 2 distinguished two dominant mass spectral patterns that distinguished the inner plaque (yellow cluster) from the rest of the tissue (orange cluster), as depicted in [Supplementary Figure S7](#). A more detailed embedding was constructed at hierarchical level 1, and it not only revealed informative structures for plaque core and outer plaque (red and green clusters, respectively) but also revealed another structure within the inner plaque (blue cluster).

The results are in concordance with those previously reported by Patterson et al.;<sup>50</sup> that is, we identified distinct molecular patterns in three main regions, namely: (1) fibrous cap (inner plaque), (2) plaque core, and (3) outer plaque (connective tissue). However, our results depict more heterogeneity within the inner and middle plaque regions. This might reflect the power of HSNE in preserving local structures of the high-dimensional spectra and thus preserves the original nonlinear manifold in the lower dimensional space.

## 4. DISCUSSION

The proposed methodology is the first of its kind, to the best of our knowledge, to handle the computational challenges of 3D MSI data analysis at full spatial and mass spectral resolution and in a reasonable time frame while maintaining high accuracy. We have shown the efficiency of this pipeline in analyzing 3D MSI data sets collected from four different biological systems and acquired by different mass spectrometers. The backbone of this methodology is HSNE, which first constructs a hierarchical representation of the high-dimensional data using the landmarks and then interactive construction of a hierarchy of t-SNE embeddings. The former assures high speed as it uses only a representative subset (i.e., landmarks) of the full data.<sup>47</sup> This reduces computational overhead while maintaining the nonlinear structure of the data, thus enabling the analysis of the millions of high-dimensional voxels encountered in 3D MSI. Interactively selecting clusters throughout the HSNE hierarchy allows the spatial structure of the 3D MSI data to be readily investigated. We demonstrate that by correlating these 3D HSNE segmentation maps with the original 3D MSI  $m/z$  features (full spatial and mass spectral resolution) the individual tissue specific features can be identified.

The presented computational pipeline has proven to be highly efficient for the spatio-chemical segmentation of 3D MSI data and the identification of associated colocalized molecular features. The segmentation maps obtained using HSNE represent regions of interest that capture and summarize molecular patterns in the high-dimensionality, spatially resolved molecular data. For the 3D MSI mouse kidney data, we have achieved much finer spatial segmentation compared with the coarser results previously reported,<sup>9</sup> which thus allowed better colocalization of ion features. In the previous analysis, the 3D data set was reduced to the molecular features retained by peak-

picking the MALDI MSI data. The linear MALDI-TOF mass spectrometer used for these measurements is characterized by its low mass resolution, often leading to broad peaks that may not be reliably peak-picked.<sup>53</sup> McDonnell et al. have reported mass spectral representations, to which peak-picking algorithms designed for linear MALDI-TOF measurements have been applied, to increase peak-picking efficiency.<sup>54</sup> Nevertheless, peak picking of linear MALDI-TOF data often leads to information loss due to inefficient peak detection. Here HSNE enabled the analysis of the complete data matrix of full spectra from all voxels, without peak picking.

For the 3D MSI data set of colorectal carcinoma, the HSNE spatial segmentation maps distinguished between the tumor and connective tissues and were found to be in close agreement with the histological images; see [Supplementary Figure S1](#). For the 3D MSI data set of mouse pancreas the HSNE analysis revealed structures consistent with the known anatomy of the pancreas; for example, that characterized by insulin ( $m/z$  5805.54) and other peptides demarcated the islets of Langerhans.<sup>10</sup> Similarly, for the 3D MSI OSCC data set, the HSNE analysis identified several molecularly distinct 3D structures, one of which was characterized by colocalized defensins, small proteins produced by neutrophil infiltration into the tumor, and was reported previously.<sup>10,55</sup> Furthermore, HSNE enables these insights to be readily attained even in data sets compromised by batch affects, spatially correlated noise, and mass spectral misalignment.

The HSNE analysis has the ability to process 3D MSI data at full spectral and full spatial resolution. The HSNE constructs scatter plots showing the distribution of the landmarks based on the similarity of their mass spectral profiles, in the full high-dimensional space. However, to construct spatially mapped HSNE structures, first a set of landmarks is selected. Here clusters of closely spaced landmarks were manually selected but could be automated by using, for example, density partitioning algorithms such as ACCENSE.<sup>36</sup>

By default, HSNE does not consider the spatial origin of each voxel's mass spectrum when analyzing the 3D MSI data. Therefore, it is not strictly required to register the sequential tissue sections into a 3D volume for the HSNE analysis. However, the image registration is highly valuable for the visualization and assessment of the 3D HSNE segmentation maps. Recent technical developments could allow the registration to be automatically performed using, for example, the t-SNE based registration pipeline presented by Abdelmoula et al.<sup>42</sup> In this previous work t-SNE was used to create a segmentation map that summarized the spatial correspondences in a tissue section's MSI data set. This segmentation map was then used to register the MSI data to a histological image of the tissue section. For 3D MSI, a similar approach can be used to coregister the MSI data sets from sequential tissue sections; namely, the global registration parameters (e.g., rotation and translation) are corrected using each tissue section's individual t-SNE segmentation map. [Supplementary Figure S8](#) shows a 3D image of the protein ion at  $m/z$  6257.9 in the mouse kidney 3D MALDI MSI data set, which is localized to the renal cortex. It can be seen that the 3D volume from the original publication<sup>10</sup> ([Figure S8a](#)) contains several discontinuities, which are due to errors during registration of the sequential tissue sections. These discontinuities could be removed after automatic t-SNE-based registration using only an Euler transform (rotation and translation)<sup>56</sup> ([Figure S8b](#)). One of the challenges facing the automated creation of 3D MSI

volumes concerns the deformations that may arise during tissue processing: The nonlinear registrations needed to correct such deformations will require geometrical constraints to preserve the original tissue shape and that could be provided by a reference such as a block-face image or an in vivo image (such as MRI) of the tissue volume before sectioning.

Other recent algorithms have also focused on alleviating the scalability issue of t-SNE, such as Largevis<sup>57</sup> and approximated-tSNE<sup>58</sup> (A-tSNE). Both algorithms focus primarily on accelerating the KNN-graph creation, a computationally very intensive step of the original t-SNE algorithm, but lack the multiscale representation of HSNE. This is an important distinction because it means HSNE is implicitly more scalable in terms of computational and memory complexity and avoids the crowded maps that result from analyzing millions of data points and that would otherwise hinder the identification of clusters.<sup>48</sup>

The ability of the HSNE to handle large volumes of high-dimensional data with reasonable computational and memory complexity makes it promising for other biological application areas that face similar computational challenges, particularly areas of neurology and cancer research. For example, HSNE holds potential for the analysis of spatially resolved omics,<sup>59</sup> especially with subcellular spatial resolution, such as those produced by array tomography,<sup>60</sup> spatial transcriptomics,<sup>61</sup> and imaging mass cytometry.<sup>62,63</sup>

## 5. CONCLUDING REMARKS

We presented a computational pipeline to analyze the volumes of 3D MSI with reasonable computational and memory complexities while maintaining accuracy at full spatial and spectral resolution. This would impact the application areas of 3D MSI as it can reveal, relatively fast and in an interactive data driven manner, multiscale molecular structures that might hold biological interest. These structures are otherwise very computationally difficult to identify using alternative pipelines.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00725.

Figure S1: Comparison between H&E images and HSNE segmentation maps of 3D DESI-MSI colorectal carcinoma data set. Figure S2: Analysis of 3D MALDI-MSI data of a mouse kidney using the HSNE. Figure S3: Spectral correlation distribution for each of the HSNE spatial segmentation maps of the mouse kidney data set. Figure S4: Visualization of the most colocalized 3D  $m/z$  features in the mouse kidney data set. Figure S5: The HSNE spatial segmentation of one of the dominant foreground structures identified by analyzing 3D MSI mouse pancreas data. Figure S6: Three tissue sections (numbers 31, 32, and 33) from the 3D MALDI-MSI OSCC data set suffer from batch effect at  $m/z$ 4956 (thymosin  $\beta$ 4 signals). Figure S7: HSNE analysis of 3D-MALDI MSI data set of atherosclerotic plaque from human carotid. Figure S8: Automatic linear alignment using Euler transform (rotation and translation) improves visualization of the reconstructed 3D image of colocalized ion feature in the renal cortex of mouse kidney. (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: B.P.F.Lelieveldt@lumc.nl. Tel: +31-(0)71-52-63935.

### ORCID

Walid M. Abdelmoula: 0000-0003-3117-7389

Liam A. McDonnell: 0000-0003-0595-9491

### Author Contributions

▽ L.A.M. and B.P.F.L. contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ REFERENCES

- (1) McDonnell, L. A.; Heeren, R. M. Imaging mass spectrometry. *Mass Spectrom. Rev.* **2007**, *26* (4), 606–643.
- (2) Caprioli, R. M.; Farmer, T. B.; Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **1997**, *69* (23), 4751–4760.
- (3) Schwamborn, K.; Caprioli, R. M. Molecular imaging by mass spectrometry—looking beyond classical histology. *Nat. Rev. Cancer* **2010**, *10* (9), 639–646.
- (4) Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 Da. *Anal. Chem.* **1988**, *60* (20), 2299–2301.
- (5) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. Protein and polymer analyses up to  $m/z$  100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2* (8), 151–153.
- (6) Amstalden van Hove, E. R.; Smith, D. F.; Heeren, R. M. A. A concise review of mass spectrometry imaging. *Journal of Chromatography A* **2010**, *1217*, 3946–3954.
- (7) Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science (Washington, DC, U. S.)* **2004**, *306* (5695), 471–473.
- (8) Sinha, T. K.; Khatib-Shahidi, S.; Yankeelov, T. E.; Mapara, K.; Ehteshami, M.; Cornett, D. S.; Dawant, B. M.; Caprioli, R. M.; Gore, J. C. Integrating spatially resolved three-dimensional MALDI IMS with in vivo magnetic resonance imaging. *Nat. Methods* **2008**, *5* (1), 57–59.
- (9) Trede, D.; Schiffler, S.; Becker, M.; Wirtz, S.; Steinhilber, K.; Strehlow, J.; Aichler, M.; Kobarg, J. H.; Oetjen, J.; Dyatlov, A.; et al. Exploring three-dimensional matrix-assisted laser desorption/ionization imaging mass spectrometry data: three-dimensional spatial segmentation of mouse kidney. *Anal. Chem.* **2012**, *84* (14), 6079–6087.
- (10) Oetjen, J.; Veselkov, K.; Watrous, J.; McKenzie, J. S.; Becker, M.; Hauberg-Lotte, L.; Kobarg, J. H.; Strittmatter, N.; Mróz, A. K.; Hoffmann, F.; et al. Benchmark datasets for 3D MALDI- and DESI-imaging mass spectrometry. *GigaScience* **2015**, *4* (1), 20.
- (11) Andersson, M.; Groseclose, M. R.; Deutch, A. Y.; Caprioli, R. M. Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. *Nat. Methods* **2008**, *5* (1), 101–108.
- (12) Giordano, S.; Morosi, L.; Veglianesi, P.; Licandro, S. A.; Frapolli, R.; Zucchetti, M.; Cappelletti, G.; Falciola, L.; Pifferi, V.; Visentin, S.; et al. 3D Mass Spectrometry Imaging Reveals a Very Heterogeneous Drug Distribution in Tumors. *Sci. Rep.* **2016**, *6* (1), 37027.
- (13) Inglese, P.; McKenzie, J. S.; Mroz, A.; Kinross, J.; Veselkov, K.; Holmes, E.; Takats, Z.; Nicholson, J. K.; Glen, R. C. Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. *Chem. Sci.* **2017**, *8*, 3500.
- (14) Eberlin, L. S.; Norton, I.; Orringer, D.; Dunn, I. F.; Liu, X.; Ide, J. L.; Jarmusch, A. K.; Ligon, K. L.; Jolesz, F. A.; Golby, A. J.; et al. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (5), 1611–1616.

- (15) Seeley, E. H.; Wilson, K. J.; Yankeelov, T. E.; Johnson, R. W.; Gore, J. C.; Caprioli, R. M.; Matrisian, L. M.; Sterling, J. A. Co-registration of multi-modality imaging allows for comprehensive analysis of tumor-induced bone disease. *Bone* **2014**, *61*, 208–216.
- (16) Jiang, L.; Chughtai, K.; Purvine, S. O.; Bhujwalla, Z. M.; Raman, V.; Paša-Tolić, L.; Heeren, R. M. A.; Glunde, K. MALDI-Mass Spectrometric Imaging Revealing Hypoxia-Driven Lipids and Proteins in a Breast Tumor Model. *Anal. Chem.* **2015**, *87* (12), 5947–5956.
- (17) Van Malderen, S. J. M.; Laforce, B.; Van Acker, T.; Nys, C.; De Rijcke, M.; De Rycke, R.; De Bruyne, M.; Boone, M.; De Schampelaere, K. A. C.; Borovinskaya, O.; et al. Three-dimensional reconstruction of the tissue-specific multi-elemental distribution within *Ceriodaphnia dubia* via multimodal registration using laser ablation ICP-mass spectrometry and X-ray spectroscopic techniques. *Anal. Chem.* **2017**, *89*, 4161.
- (18) Hinsenkamp, I.; Schulz, S.; Roscher, M.; Suhr, A. M.; Meyer, B.; Munteanu, B.; Fuchser, J.; Schoenberg, S. O.; Ebert, M. P. A.; Wängler, B.; et al. Inhibition of Rho-Associated Kinase 1/2 Attenuates Tumor Growth in Murine Gastric Cancer. *Neoplasia (United States)* **2016**, *18* (8), 500–511.
- (19) Calligaris, D.; Norton, I.; Feldman, D. R.; Ide, J. L.; Dunn, I. F.; Eberlin, L. S.; Graham Cooks, R.; Jolesz, F. A.; Golby, A. J.; Santagata, S.; et al. Mass spectrometry imaging as a tool for surgical decision-making. *J. Mass Spectrom.* **2013**, *48* (11), 1178–1187.
- (20) Santagata, S.; Eberlin, L. S.; Norton, I.; Calligaris, D.; Feldman, D. R.; Ide, J. L.; Liu, X.; Wiley, J. S.; Vestal, M. L.; Ramkissoon, S. H.; et al. Intraoperative mass spectrometry mapping of an onco-metabolite to guide brain tumor surgery. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (30), 11121–11126.
- (21) Thiele, H.; Heldmann, S.; Trede, D.; Strehlow, J.; Wirtz, S.; Dreher, W.; Berger, J.; Oetjen, J.; Kobarg, J. H.; Fischer, B.; et al. 2D and 3D MALDI-imaging: Conceptual strategies for visualization and data mining. *Biochim. Biophys. Acta, Proteins Proteomics* **2014**, *1844*, 117–137.
- (22) Ogrinc Potočnik, N.; Porta, T.; Becker, M.; Heeren, R. M. A.; Ellis, S. R. Use of advantageous, volatile matrices enabled by next-generation high-speed matrix-assisted laser desorption/ionization time-of-flight imaging employing a scanning laser beam. *Rapid Commun. Mass Spectrom.* **2015**, *29* (23), 2195–2203.
- (23) Steven, R. T.; Dexter, A.; Bunch, J. Investigating MALDI MSI parameters (Part 1) - A systematic survey of the effects of repetition rates up to 20kHz in continuous raster mode. *Methods* **2016**, *104*, 101–110.
- (24) Palmer, A. D.; Alexandrov, T. Serial 3D imaging mass spectrometry at its tipping point. *Anal. Chem.* **2015**, *87* (8), 4055–4062.
- (25) Trede, D.; Schiffler, S.; Becker, M.; Wirtz, S.; Steinhorst, K.; Strehlow, J.; Aichler, M.; Kobarg, J. H.; Oetjen, J.; Dyatlov, A.; et al. Exploring three-dimensional matrix-assisted laser desorption/ionization imaging mass spectrometry data: Three-dimensional spatial segmentation of mouse kidney. *Anal. Chem.* **2012**, *84* (14), 6079–6087.
- (26) Dexter, A.; Race, A. M.; Steven, R. T.; Barnes, J. R.; Hulme, H.; Goodwin, R. J. A.; Styles, I. B.; Bunch, J. Two-Phase and Graph-Based Clustering Methods for Accurate and Efficient Segmentation of Large Mass Spectrometry Images. *Anal. Chem.* **2017**, *89* (21), 11293–11300.
- (27) Van Der Maaten, L. J. P.; Postma, E. O.; Van Den Herik, H. J. Dimensionality Reduction: A Comparative Review. *J. Mach. Learn. Res.* **2009**, *10*, 1–41.
- (28) Fukunaga, K. Introduction to statistical pattern recognition. *Pattern Recognit.* **1990**, *22* (7), 833–834.
- (29) Thomas, S. A.; Race, A. M.; Steven, R. T.; Gilmore, I. S.; Bunch, J. Dimensionality Reduction of Mass Spectrometry Imaging Data using Autoencoders. *IEEE Symp. Ser. Comput. Intell.* **2016**, 1–7.
- (30) Jolliffe, I. T. Principal Component Analysis and Factor Analysis. In *Principal Component Analysis*; Springer, 1986; pp 115–128.
- (31) Lee, D. D.; Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401* (6755), 788–791.
- (32) Jones, E. A.; van Remoortere, A.; van Zeijl, R. J. M.; Hogendoorn, P. C. W.; Bovée, J. V. M. G.; Deelder, A. M.; McDonnell, L. A. Multiple statistical analysis techniques corroborate intratumor heterogeneity in imaging mass spectrometry datasets of myxofibrosarcoma. *PLoS One* **2011**, *6* (9), e24913.
- (33) Veselkov, K. A.; Mirnezami, R.; Strittmatter, N.; Goldin, R. D.; Kinross, J.; Speller, A. V. M.; Abramov, T.; Jones, E. A.; Darzi, A.; Holmes, E.; et al. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (3), 1216–1221.
- (34) Race, A. M.; Steven, R. T.; Palmer, A. D.; Styles, I. B.; Bunch, J. Memory efficient principal component analysis for the dimensionality reduction of large mass spectrometry imaging data sets. *Anal. Chem.* **2013**, *85* (6), 3071–3078.
- (35) Jones, E. A.; Shyti, R.; van Zeijl, R. J. M.; van Heiningen, S. H.; Ferrari, M. D.; Deelder, A. M.; Tolner, E. A.; van den Maagdenberg, A. M. J. M.; McDonnell, L. A. Imaging mass spectrometry to visualize biomolecule distributions in mouse brain tissue following hemispheric cortical spreading depression. *J. Proteomics* **2012**, *75* (16), S027–S035.
- (36) Shekhar, K.; Brodin, P.; Davis, M. M.; Chakraborty, A. K. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (1), 202–207.
- (37) Hinton, G. E.; Roweis, S. T. Stochastic neighbor embedding. *NIPS* **2002**, 833–840.
- (38) van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
- (39) Ji, S. Computational genetic neuroanatomy of the developing mouse brain: dimensionality reduction, visualization, and clustering. *BMC Bioinf.* **2013**, *14*, 222.
- (40) Huisman, S. M. H.; Van Lew, B.; Mahfouz, A.; Pezzotti, N.; Hollt, T.; Michielsen, L.; Vilanova, A.; Reinders, M. J. T.; Lelieveldt, B. P. F. BrainScope: Interactive visual exploration of the spatial and temporal human brain transcriptome. *Nucleic Acids Res.* **2017**, *45* (10), e83.
- (41) Fonville, J. M.; Carter, C. L.; Pizarro, L.; Steven, R. T.; Palmer, A. D.; Griffiths, R. L.; Lalor, P. F.; Lindon, J. C.; Nicholson, J. K.; Holmes, E.; et al. Hyperspectral visualization of mass spectrometry imaging data. *Anal. Chem.* **2013**, *85* (3), 1415–1423.
- (42) Abdelmoula, W. M.; Skraskova, K.; Balluff, B.; Carreira, R. J.; Tolner, E. A.; Lelieveldt, B. P.; van der Maaten, L.; Morreau, H.; van den Maagdenberg, A. M.; Heeren, R. M.; et al. Automatic generic registration of mass spectrometry imaging data to histology using nonlinear stochastic embedding. *Anal. Chem.* **2014**, *86* (18), 9204–9211.
- (43) van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* **2014**, *15*, 3221–3245.
- (44) Mahfouz, A.; van de Giessen, M.; van der Maaten, L.; Huisman, S.; Reinders, M.; Hawrylycz, M. J.; Lelieveldt, B. P. F. Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. *Methods* **2015**, *73*, 79–89.
- (45) Abdelmoula, W. M.; Balluff, B.; Englert, S.; Dijkstra, J.; Reinders, M. J. T.; Walch, A.; McDonnell, L. A.; Lelieveldt, B. P. F. Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of mass spectrometry imaging data. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (43), 12244–12249.
- (46) van Unen, V.; Li, N.; Molendijk, I.; Temurhan, M.; Höllt, T.; van der Meulen-de Jong, A. E.; Verspaget, H. W.; Mearin, M. L.; Mulder, C. J.; van Bergen, J.; et al. Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets. *Immunity* **2016**, *44* (5), 1227–1239.
- (47) Pezzotti, N.; Höllt, T.; Lelieveldt, B.; Eisemann, E.; Vilanova, A. Hierarchical Stochastic Neighbor Embedding. *Comput. Graph. Forum* **2016**, *35* (3), 21–30.
- (48) van Unen, V.; Höllt, T.; Pezzotti, N.; Li, N.; Reinders, M. J. T.; Eisemann, E.; Koning, F.; Vilanova, A.; Lelieveldt, B. P. F. Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. *Nat. Commun.* **2017**, *8* (1), 1740.

(49) Shneiderman, B. The eyes have it: a task by data type taxonomy for information visualizations. *Proc. 1996 IEEE Symp. Vis. Lang.* **1996**, 336–343.

(50) Patterson, N. H.; Doonan, R. J.; Daskalopoulou, S. S.; Dufresne, M.; Lenglet, S.; Montecucco, F.; Thomas, A.; Chaurand, P. Three-dimensional imaging MS of lipids in atherosclerotic plaques: Open-source methods for reconstruction and analysis. *Proteomics* **2016**, *16* (11–12), 1642–1651.

(51) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.* **1979**, *28* (1), 100.

(52) Bezdek, J. C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10* (2–3), 191–203.

(53) Mantini, D.; Petrucci, F.; Pieragostino, D.; Del Boccio, P.; Di Nicola, M.; Di Ilio, C.; Federici, G.; Sacchetta, P.; Comani, S.; Urbani, A. LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinf.* **2007**, *8* (1), 101.

(54) McDonnell, L. A.; van Remoortere, A.; de Velde, N.; van Zeijl, R. J.; Deelder, A. M. Imaging mass spectrometry data reduction: automated feature identification and extraction. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1969–1978.

(55) Cheng, C. C.; Chang, J.; Chen, L. Y.; Ho, A. S.; Huang, K. J.; Lee, S. C.; Mai, F.-D.; Chang, C. C. Human neutrophil peptides 1–3 as gastric cancer tissue markers measured by MALDI-imaging mass spectrometry: Implications for infiltrated neutrophils as a tumor target. *Dis. Markers* **2012**, *32* (1), 21–31.

(56) Klein, S.; Staring, M.; Murphy, K.; Viergever, M. A.; Pluim, J. P. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **2010**, *29* (1), 196–205.

(57) Tang, J.; Liu, J.; Zhang, M.; Mei, Q. Visualization Large-Scale and High-Dimensional Data. 2016, arXiv:1602.00370. arXiv.org e-Print archive. <https://arxiv.org/abs/1602.00370>.

(58) Pezzotti, N.; Lelieveldt, B. P. F.; van Der Maaten, L.; Höllt, T.; Eisemann, E.; Vilanova, A. Approximated and user steerable tSNE for progressive visual analytics. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23* (7), 1739–1752.

(59) Crosetto, N.; Bienko, M.; van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **2015**, *16* (1), 57–66.

(60) Micheva, K. D.; Smith, S. J. Array Tomography: A New Tool for Imaging the Molecular Architecture and Ultrastructure of Neural Circuits. *Neuron* **2007**, *55* (1), 25–36.

(61) Ståhl, P. L.; Salmén, F.; Vickovic, S.; Lundmark, A.; Navarro, J. F.; Magnusson, J.; Giacomello, S.; Asp, M.; Westholm, J. O.; Huss, M.; et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (Washington, DC, U. S.)* **2016**, *353* (6294), 78–82.

(62) Giesen, C.; Wang, H. A. O.; Schapiro, D.; Zivanovic, N.; Jacobs, A.; Hattendorf, B.; Schüffler, P. J.; Grolimund, D.; Buhmann, J. M.; Brandt, S.; et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **2014**, *11* (4), 417–422.

(63) Chang, Q.; Ornatsky, O. I.; Siddiqui, I.; Loboda, A.; Baranov, V. I.; Hedley, D. W. Imaging Mass Cytometry. *Cytometry, Part A* **2017**, *91*, 160–169.